IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SPEECH-BASED DETECTION OF PARKINSON'S DISEASE USING SPECTRAL FEATURES AND RANDOM FOREST CLASSIFIER

Mrunal Gholap , Vishal Aher, Satish Turkane

Department of E&TC Engineering, Department of E&TC Engineering, Department of E&TC Engineering,

Pravara Rural Engineering College, Loni, India.

Abstract: Parkinson disease (PD) is a progressive neurodegenerative disease, which has a significant impact on speech production due to motor disabilities. The early and precise diagnosis of PD is a burning clinical issue that indicates the necessity of non-invasive and valid diagnostic methods. This paper proposes a speech-based method of PD detection, and it employs spectral acoustic features and a support vector machine (SVM) classifier. Spectral descriptors such as Mel Frequency Cepstral Coefficients (MFCC) and Mel-spectrogram coefficients are received by acoustic signals in order to capture pathologic variations in the vocal patterns. Experimental findings show that the suggested framework provides an accuracy of 85% exceeding a number of the state-of-the-art approaches. These results suggest that spectral speech biomarkers together with machine learning algorithms provide an effective and non-invasive method of screening PD. Therefore, the suggested method has great potential in aiding the non-invasive diagnosis in the clinical environment.

Index Terms – PD, Speech signal, RF ,Spectral Feature.

I. INTRODUCTION

Parkinson disease (PD) is a chronic neurodegenerative disease, which causes gradual worsening of motor activity. It is a disorder that is caused by the selective loss or dysfunction of the dopaminergic neurons in the brain, especially the neurons that are involved in the manufacture of dopamine. One of the neurotransmitters of movement and coordination, dopamine is deficient, leading to the typical motor symptoms of PD, namely tremor, bradykinesia, postural instability, and akinesia, and speech and handwriting changes. As demonstrated in clinical evidence, patients with PD often have respiratory and swallowing issues, which, in turn, affect different speech aspects of these individuals [1]. These patients have a loss of vocal control, this is in the form of slurred, breathy, hoarse and low-volume speech.

Speech feature extraction is a crucial element of many applications such as voice recognition, speaker recognition, emotion recognition and speech pathology evaluation. It consists in the recognition and extraction of significant features, or patterns, of raw audio signals. These characteristics extract critical information of the vocal stream, which is the pitch, intensity, spectral qualities, and time fluctuations. The most commonly used methods are Mel-frequency cepstral coefficients (MFCCs) that represent the speech frequency properties in a similar fashion as the human auditory system, and pitch-identification algorithms that determine the underlying frequency of spoken words as fundamental frequency bands in speech synthesis. The obtained characteristics are then used as inputs in various speech-processing pipelines. Over the past few years, various methodologies of feature extraction have been developed in order to model various speech deficits. Numerous researches have been conducted on patients with the PD by studying their speech cues. PD evaluation through

speech essentially depends on the acoustic, spectral and cepstral features. After the extraction of features, a suitable machine-learning method is implemented to test the discriminative capability of features retrieved in the identification of PD. Sakar et al. [2] created a massive data pool of speech samples of the subjects within the PD and normal control groups. The database has the recording of sustained vowels, phrases and words. It was measured by the extraction of diverse acoustic characteristics. Little et al. used a smartphone application to record the voice of patients with PD during their daily routines [3]. The researchers also tested the features pertaining to vocal tremor and dysphonia, which proved the possibility of using mobile technology to monitor the PD symptoms on a long-term basis. Gomez ´ et al. [4] examined the utilization of articulatory kinematic features derived from speech, investigating the correlation between speech production mechanisms and PD. Kinematic analysis approaches were utilized to evaluate speech motor abnormalities in patients with PD. Warulet al. [5] utilized chirplet transform for the classification of PD. Hires et al. [6] ˇ presented a CNN ensemble trained with various fine-tuning methodologies for the categorization of PD and HC. Warulet al. [7] employed wavelet synchrosqueezing transform for the classification of PD. Karan et al. [8] utilized features obtained from the Hilbert transform alongside an SVM classifier for PD classification. Vasquez-Correa et al. [9] employed feature representations derived from deep autoencoders for the categorization of PD. Garcia et al. [10] concentrated on advanced linguistic attributes, encompassing semantic categories, grammatical frameworks, and recurrence patterns. Narendra and Alku [11] investigated glottal source attributes related to vocal fold dynamics for PD categorization. Garcia et al. [12] employed articulatory, prosodic, and phonemic features for PD categorization.

The effectiveness of a speech categorization system is significantly affected by the careful selection of discriminative features and robust classifiers. Researchers investigated into a lot of different ways to extract features and machine learning classifiers for finding PD using speech signals. In this study, we systematically examine the efficacy of diverse spectral features in differentiating between PD-affected and healthy speech. To derive insightful acoustic information from speech recordings, we extracted spectral features like MFCC, perceptual linear prediction coefficients (PLPC), spectral contrast, spectral bandwidth, spectral centroid, spectral flatness, and spectral roll-off. The PD affects the motor control system, which affect clarity and change the stability of the voice. These physiological and neuromuscular changes affect vocal folds vibratation and the shape of the vocal tract, which changes the spectral structure and energy distribution of the speech signal. Spectral analysis allows for the measurement of these variations in the frequency domain. Consequently, it is believed that spectral features offer improved discriminative capacity for the effective classification of PD-affected and healthy speech conditions. The structure of this paper is as follows: The PC-GITA database, which was used to assess the proposed system, is covered in Section II. The proposed strategy for using speech signals to diagnose PD is covered in Section III. The results of the experiment and a discussion are presented in Section IV. Section V presents the conclusion

II. PC-GITA DATABASE

This study employs the PC-GITA database [1] to assess the proposed framework. The dataset includes audio recordings from 50 individuals diagnosed with PD and 50 healthy subjects. The database shows that the gender and age demographics are evenly split. Men with PD are between the ages of 33 and 81, and women with PD are between the ages of 49 and 75. For healthy people, the age ranges are 31 to 86 for men and 43 to 76 for women. The data was sampled at 44.1 kHz with a resolution of 16 bits. This study analyzes the phonation characteristics of five prolonged Spanish vowels, utilizing data from the PC-GITA database. This recording includes three repetitions of each vowel (/a/, /e/, /i/, /o/, and /u/) from both PD patients and healthy individuals, for a total of 300 recordings per vowel.

III. CLASSIFICATION METHOD

The block diagram of the proposed framework for detecting PD from speech is shown in Fig. 1. The raw speech signal is first preprocessed and spectral features are extracted from it to get unique acoustic properties. Finally, a random forest (RF) classifier uses these features to categorize the speech into PD and healthy groups.

A. Pre-processing involves a number of steps, such as preemphasis, the removal of silence, and framing [13], [14]. A pre-emphasis filter is used to improve the speech signal at first. The basic purpose of a pre-emphasis filter is to enhance higher frequencies by boosting their amplitudes relative to those of lower frequencies. The pre-emphasis filter is a type of high-pass filter that makes the spectrum of a signal equal by selectively boosting the higher frequency parts. Short-term energy analysis is then used to find the active speech

segments by removing the silent parts of the emphasized speech. In the end, active speech segments are split into many frames, each lasting 20 milliseconds and overlapping by 10 milliseconds.

B. Feature extraction Feature extraction is the process of reducing the number of dimensions in data while keeping the important information that can be used for voice classification tasks. In this study, we extracted various spectral properties from the speech signal for PD classification.

1) Mel-frequency cepstral coefficients (MFCC): The MFCC is the most common spectral features used in audio and speech processing. These features are designed to represent the short-term power spectrum of a speech signal while incorporating perceptual characteristics of the human auditory system [15], [16], [17], [18]. The Mel scale is used to extract MFCCs. It models how the human ear is more sensitive to lower frequencies than higher ones, which gives lower frequency components more detail. The MFCCs give a compact representation that keeps important sound information and gets rid of information that isn't needed. They do this by transforming the speech spectrum into the Mel domain and employing cepstral analysis. This study calculates 13 static MFCC coefficients and their first-order (Δ) and secondorder (Δ-Δ) derivatives for each speech frame, resulting in a 39-dimensional feature vector that represents both spectral structure and temporal dynamics.

2) Perceptual Linear Prediction coefficients (PLPC): The PLPC are spectral features to model the human auditory system more closely than conventional cepstral representations . Hermansky introduced PLPC [19], [20] to use psychoacoustic ideas like critical-band spectral resolution, equal-loudness preemphasis, and the intensity–loudness power law in the process of extracting features. In PLPC analysis, the short-term power spectrum of the speech signal is initially transformed to the Bark scale to simulate the frequency resolution of human hearing. Then, an equal-loudness curve is used to weight the spectrum, and a cubic-root intensity law is used to compress it to mimic how loudness is perceived in a nonlinear way. Finally, linear prediction analysis is used on the changed spectrum to get coefficients that are similar to cepstral coefficients. The PLPC are better than MFCC at handling noise and changes in the spectrum while keeping information that is important to perception. The PLPC features are especially helpful for finding PD because they focus on vocal tract resonances and spectral envelope characteristics that may change because of hypokinetic dysarthria, less precise articulation, and phonatory instability. The PLPC can improve class separability between healthy and PD subjects by capturing perceptually meaningful spectral distortions in speech affected by PD. Consequently, integrating PLP features with traditional spectral descriptors may enhance the discriminative power of the proposed speech based diagnostic framework. We have extracted 13 PLPC features for classification of PD.

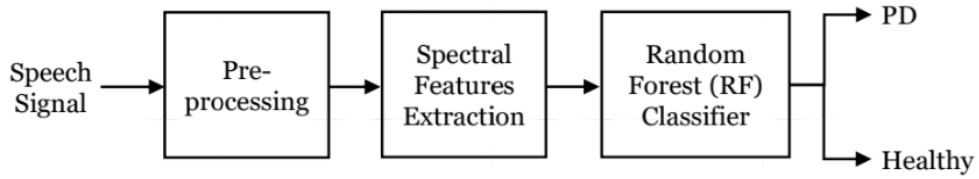

Fig. 1: The proposed speech signal based method for classifying PD and HC

3) Spectral contrast: Spectral contrast is a frequency domain feature demonstrating energy difference between spectral peaks and valleys in certain parts of the speech spectrum. Cepstral features mostly show the overall spectral envelope, while spectral contrast focuses on the relative distribution of harmonic and non-harmonic components across different frequency ranges [21], [22]. First spectrum is divided into different frequency bands and find the logarithmic difference between the highest (peak) and lowest (valley) energy levels in each band. The illustration shows changes in spectral dynamics and resonance patterns very well. For speech analysis for PD, spectral contrast is very important because neuromotor problems can change the harmonic structure, make articulation less clear, and change how energy is spread across frequency bands. So, spectral contrast features can find small spectral irregularities that are connected to bad phonation.

4) Spectral centroid: The spectral centroid is a common frequency-domain descriptor that shows the center of mass of the magnitude spectrum [23], [24]. The weighted mean of the frequency components is used to find it, with the weights being the size of the spectrum at each frequency bin. The spectral centroid shows where most of the spectral energy is concentrated. It is often connected to the sound signal's brightness perceptual attribute. When the centroid value is higher, it means that more energy is spread out over higher frequencies. When the value is lower, it means that low-frequency components are more important. Changes

in vocal fold vibration, less precise articulation, and changes in resonance characteristics can all change the distribution of spectral energy when analyzing PD speech. Because of this, spectral centroid can be a useful feature for finding frequency shifts and energy redistribution patterns in speech that is affected by PD. This can help improve classification performance.

5) Spectral bandwidth: The spectral bandwidth is a frequency domain property that quantifies spectral dispersion around its center [23], [24]. It gives you an idea of how much the energy of a speech signal is distributed in the frequency spectrum. In the event that the bandwidth is small then most of the energy lies in the middle. At a bigger bandwidth, the spectrum is more changeable and has a bigger frequency range. Issues of neuromotor may alter how speech sounds occur, by influencing the speech sound production by the mouth and the voice stability. This has the ability to alter the resonance properties and spectral power. Therefore, spectral bandwidth is capable of detecting changes in frequency diffusing related to pathological speech production making it an important quality of classifying. Spectral bandwidth is a property in the frequency domain that measures spread of the spectrum from its center [23], [24]. In other words, spectral bandwidth tells you how much the energy of a speech signal is spread out across the frequency spectrum. When the bandwidth is small, the energy is mostly in the middle. When the bandwidth is larger, the spectrum is more variable and has a wider range of frequencies. Neuromotor problems can change the way speech sounds by affecting the way the mouth moves and the stability of the voice. This can change the resonance characteristics and spectral energy distribution. Thus, spectral bandwidth can identify variations in frequency spread associated with pathological speech production, rendering it a significant attribute for classification.

6) Spectral flatness: The degree of flatness in a spectrum is evaluated by the spectral flatness measure [25]. It gauges how evenly the power spectrum's frequency distribution is distributed. It is usually found by dividing the geometric mean of the power spectrum in a frame by the arithmetic mean of the power spectrum in that frame. A higher spectral flatness value means that the energy is spread out more evenly across frequencies, which is typical of noise-like signals. A lower value, on the other hand, means that there are strong harmonic peaks. In the context of speech analysis for PD, pathological phonation and irregular vocal fold vibration may make speech sound more breathy or less regular, which can change the harmonic structure of speech. Spectral flatness is a useful tool for measuring changes in spectral regularity, which makes it a useful feature for picking up on small changes in voice quality that are linked to PD

7) Spectral roll-off: Spectral roll-off is a feature in the frequency domain that shows the frequency below which a certain percentage usually 85% or 95% of the total spectral energy is concentrated [23]. It gives an estimate of the highest frequency that the main energy distribution in a speech signal can reach. A higher roll-off frequency means that more energy is in the higher frequency range, while a lower value means that more energy is in the lower frequency range. In PD speech analysis, motor dysfunctions and modified vocal fold dynamics can affect the harmonic structure and energy distribution of speech. Pathological changes like these may change the concentration of spectral energy, which can be accurately measured by spectral roll-off. This feature allows to capture the information for classification of PD. C. Random Forest (RF) The RF classifier is a widely utilized machine learning technique that is predominantly employed for jobs involving classification [17]. It is ensemble learning that combines the predictions of many decision trees to make the final predictions more accurate and reliable. The RF are widely used in many fields, such as image classification, fraud detection, and disease diagnosis. You make the RF by putting together a group of decision trees. To build each decision tree, a random selection of the training data and features is used. Adding randomness helps to fix the problem of overfitting. The RF algorithm combines the predictions made by different trees into one by using a majority voting system for classification tasks or by averaging for regression tasks. As a result, this process produces a definitive prediction that is often more accurate and reliable than that of a single decision tree.

IV. RESULTS AND DISCUSSION

In this section, a critical analysis of the discriminating power of the proposed spectral features on RF classifiers is elaborated. The classification tests were based on sustained vowel speech tasks that were obtained in the PC-GITA database. The evaluation was done using a ten-fold cross-validation procedure to ensure that it is robust and general. Table I results indicate spectral features classification using RF classifier of tasks on sustained vowel of PC-GITA database. The findings are reported in the accuracy, F1-score, preciseness, and ROCAUC that are reported in mean ± standard deviation, using ten-fold cross-validation. Vowel /e/ had highest classification accuracy of 0.907 ± 0.047 and good F1-score of 0.906 ± 0.055, and this implies that it was performing well in recall and precision. The vowel /i/ also worked well with an accuracy of 0.897 ± 0.053 and the maximum ROC-AUC of 0.964 ± 0.023. The same happened with the vowels /a/ and /o/ with the accuracy of both being 0.880 and the standard deviation of both being rather consistent showing similar classification accuracy among folds. Conversely, vowel /u/ was least accurate 0.840 +/-

0.057 and less precise (0.796 ± 0.078) which was a pointer of the low discriminative ability of this vowel. Table II provides a comparison of the performance metrics associated with various existing methods alongside the proposed methodology for detecting PD utilizing speech signals. The table presents the types of features extracted, the classification algorithms employed, and the corresponding classification accuracy achieved for each method. Vasquez-Correa et al. [9] utilized autoencoder-based features and CNN classifier, achieving an accuracy of 84%. Karan et al. [8] utilized features based on the Hilbert transform to attain 90% accuracy. Garcia et al. [10] concentrated on linguistic attributes, including semantic fields, grammatical features, and word repetition, attaining a comparatively lower accuracy of 66% with SVM. Narendra and Alku [11] utilized glottal features, associated with vocal fold dynamics, achieving an accuracy of 69.55%. Garcia et al. [12] employed the various articulatory, prosodic, and honemic features to get 84% accuracy. The proposed method, which utilizes spectral features achieves an accuracy of 90%. This performance demonstrates the effectiveness of spectral features in capturing the acoustic markers of PD and shows competitive results when compared to existing approaches

TABLE I: Classification results on sustained vowels using spectral features and RF classifier

Task	Accuracy	F1-score	Precision	ROC AUC
Vowel /a/	0.880 ± 0.065	0.881 ± 0.068	0.873 ± 0.093	0.957 ± 0.028
Vowel /e/	0.907 ± 0.047	0.906 ± 0.055	0.883 ± 0.065	0.957 ± 0.028
Vowel /i/	0.897 ± 0.053	0.897 ± 0.054	0.863 ± 0.072	0.957 ± 0.028
Vowel /o/	0.880 ± 0.050	0.884 ± 0.048	0.857 ± 0.077	0.957 ± 0.028
Vowel /u/	0.840 ± 0.057	0.849 ± 0.058	0.796 ± 0.078	0.928 ± 0.058

Summarizes a comparative evaluation between previously reported PD speech classification approaches and the proposed spectral feature-based framework. The comparison encompasses the utilized acoustic or linguistic features, the selected classification model, and the associated classification accuracy. Karan et al. [8] utilized features obtained from the Hilbert transform, attaining an enhanced accuracy of 90% with an SVM classifier. Vasquez-Correa et al. [9] employed deep autoencoder-based feature representations alongside a CNN, achieving an accuracy of 84%. Conversely, Garcia et al. [10] concentrated on advanced linguistic attributes, such as semantic categories, grammatical structure, and repetition patterns, yielding a relatively lower performance (66%). Narendra and Alku [11] examined glottal source characteristics associated with vocal fold dynamics and reported an accuracy of 69.55%. Garcia et al. [12] also used articulatory, prosodic, and phonemic descriptors, which led to an 84% accuracy rate in classification. The proposed structure uses a wide range of spectral features, such as MFCC, chromagram, and other spectral descriptors, along with a RF classifier. The 90% accuracy shows that spectral-domain representations can accurately show acoustic changes that are linked to PD. The results indicate that the proposed method provides competitive performance compared to existing techniques while maintaining a relatively simple and computationally efficient feature.extraction strategy. The diagnosis of the PD in a clinical setting is quite difficult, and the period of time necessary to come to the final diagnosis may span up to two years or more. At the moment, there is no single diagnostic test that is specific and therefore can be used conclusively to ascertain the presence of PD. Therefore, a detailed clinical assessment is the main tool of neurologists when reaching a diagnosis, which includes analysis of the medical history of a patient, observation of typical motor and non-motor symptoms, and performance of a complete neurological examination. In comparison, the current study aims at obtaining discriminative spectral representations directly out of the speech samples with a view to identify the pathological variations related to PD. The spectral properties that are extracted are then used as inputs in machine-learning classifiers to detect PD automatically.

V. CONCLUSION

.

In this research, we have introduced a spectral-feature based framework of automated detection of the PD on sustained vowel utterances. A large set of spectral descriptors were used in order to reflect pathological acoustic changes related to PD. Classification performance was measured using RF classifiers that were trained on the PC-GITA database. The results of experiments prove that the chosen spectral features are effective in distinguishing the healthy and PD speech signals, and the highest classification rates are achieved in all sustained vowels. Some vowels showed significantly better separability, which indicates the

role of articulatory structure and resonant quality in the spectral appearance of PD. The high values of ROC-AUC are also consistent with the ability of the proposed spectral representation to represent clinically significant acoustic differences. These findings support the validity of spectral-domain acoustic biomarkers as a useful, non-invasive screening modality of PD. Future research will be aimed at adding perceptually driven and nonlinear spectral characteristic, implementing the enhanced feature-selection methods, and testing the framework on bigger and multilingual speech data to improve its generalizability and clinical usefulness.

VI REFERENCES

[1] J. R. Orozco-Arroyave, J. D. Arias-Londono, J. F. Vargas- ˜ Bonilla, M. C. Gonzalez-Rativa, and E. N ´ oth, "New ¨ spanish speech corpus database for the analysis of people suffering from parkinson's disease," in Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), 2014, pp. 342–347.

[2] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a parkinson speech dataset with multiple types of sound recordings," IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 4, pp. 828–834, 2013.

[3] M. Little, P. McSharry, E. Hunter, J. Spielman, and L. Ramig, "Suitability of dysphonia measurements for telemonitoring of parkinson's disease," Nature Precedings, pp. 1–1, 2008.

[4] P. G. Vilda, J. Mekyska, A. G. Rodellar, D. P. Alonso, V. R. Biarge, and A. A. Marquina, "Monitoring parkinson disease from speech articulation kinematics," Loquens: revista espanola de ciencias del habla , no. 4, p. 2, 2017

[5] P. Warule, S. P. Mishra, and S. Deb, "Time-frequency analysis of speech signal using chirplet transform for automatic diagnosis of parkinson's disease," Biomedical Engineering Letters, pp. 1–11, 2023.

[6] M. Hires, M. Gazda, P. Drot ˇ ar, N. D. Pah, M. A. ´ Motin, and D. K. Kumar, "Convolutional neural network ensemble for parkinson's disease detection from voice recordings," Computers in biology and medicine, vol. 141, p. 105021, 2022.

[7] P. Warule, S. P. Mishra, and S. Deb, "Time-frequency analysis of speech signal using wavelet synchrosqueezing transform for automatic detection of parkinson's disease," IEEE Sensors Letters, 2023.

[8] B. Karan, S. S. Sahu, J. R. Orozco-Arroyave, and K. Mahto, "Hilbert spectrum analysis for automatic detection and evaluation of parkinson's speech," Biomedical Signal Processing and Control, vol. 61, p. 102050, 2020.

[9] J. C. Vasquez-Correa, T. Arias-Vergara, M. Schuster, J. R. Orozco-Arroyave, and E. Noth, "Parallel representation ¨ learning for the classification of pathological speech: studies on parkinson's disease and cleft lip and palate," Speech Communication, vol. 122, pp. 56–67, 2020.

[10] A. M. Garc´ıa, F. Carrillo, J. R. Orozco-Arroyave, N. Trujillo, J. F. V. Bonilla, S. Fittipaldi, F. Adolfi, E. Noth, ¨ M. Sigman, D. F. Slezak et al., "How language flows when movements don't: an automated analysis of spontaneous discourse in parkinson's disease," Brain and language, vol. 162, pp. 19–28, 2016

.[11] N. Narendra and P. Alku, "Automatic assessment of intelligibility in speakers with dysarthria from coded telephone speech using glottal features," Computer Speech & Language, vol. 65, p. 101117, 2021.

[12] A. M. Garc´ıa, T. Arias-Vergara, J. C Vasquez-Correa, E. Noth, M. Schuster, A. E. Welch, Y. Bocanegra, ¨ A. Baena, and J. R. Orozco-Arroyave, "Cognitive determinants of dysarthria in parkinson's disease: an automated machine learning approach," Movement Disorders, vol. 36, no. 12, pp. 2862–2873, 2021.

[13] S. P. Mishra, P. Warule, and S. Deb, "Fixed frequency range empirical wavelet transform based acoustic and entropy features for speech emotion recognition," Speech Communication, vol. 166, p. 103148, 2025.

[14] P. Warule, S. Chandratre, S. Daware, S. P. Mishra, and S. Deb, "Dual-tree complex wavelet transform for the automatic detection of the common cold based on speech signals," Circuits, Systems, and Signal Processing, vol. 44, no. 7, pp. 5107–5126, 2025.

[15] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 357–366, 1980.

[16] S. P. Mishra, P. Warule, and S. Deb, "Speech emotion recognition using mfcc-based entropy feature," Signal, Image and Video Processing, pp. 1–9, 2023.

[17] S. S. Nayak, A. D. Darji, and P. K. Shah, "Machine learning approach for detecting covid-19 from speech signal using mel frequency magnitude coefficient," Signal, Image and Video Processing, pp. 1–8, 2023.

[18] P. Warule, S. P. Mishra, S. Deb, and D. Joshi, "Empirical mode decomposition based detection of common cold using speech signal," in TENCON 2023-2023 IEEE Region 10 Conference (TENCON). IEEE, 2023, pp. 899–903.
[19] H. Hermansky, "Perceptual linear predictive (plp) analysis of speech," the Journal of the Acoustical Society of America, vol. 87, no. 4, pp. 1738–1752, 1990.
[20] P. Warule, S. P. Mishra, S. Deb, and J. Krajewski, "Sinusoidal model-based diagnosis of the common cold from the speech signal," Biomedical Signal Processing and Control, vol. 83, p. 104653, 2023.
[21] D.-N. Jiang, L. Lu, H.-J. Zhang, J.-H. Tao, and L.- H. Cai, "Music type classification by spectral contrast feature," in Proceedings. IEEE international conference on multimedia and expo, vol. 1. IEEE, 2002, pp. 113–116.
[22] P. Warule, S. Chandratre, S. P. Mishra, and S. Deb, "Detection of the common cold from speech signals using transformer model and spectral features," Biomedical Signal Processing and Control, vol. 93, p. 106158, 2024.
[23] M. Aly, K. H. Rahouma, and S. M. Ramzy, "Pay attention to the speech: Covid-19 diagnosis using machine learning and crowdsourced respiratory and speech recordings," Alexandria Engineering Journal, vol. 61, no. 5, pp. 3487–3500, 2022.
[24] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the cuidado project," CUIDADO Ist Project Report, vol. 54, no. 0, pp. 1–25, 2004.
[25] A. Ramalingam and S. Krishnan, "Gaussian mixture modeling of short-time fourier transform features for audio fingerprinting," IEEE Transactions on Information Forensics and Security, vol. 1, no. 4, pp. 457–463, 2006.

.